R E P O R T R E S U M E S

ED 015 118 24 RE 001 007

SELECTED READING READINESS TESTS AS PREDICTORS OF SUCCESS IN READING.

BY- LOWELL, ROBERT E.

REPORT NUMBER BR-6-8894 PUB DATE 31 JUL 67
GRANT OEG-1-7-068894-0323
EDRS PRICE MF-$0.25 HC-$1.52 36P.

DESCRIPTORS- *READING READINESS TESTS, *PROGNOSTIC TESTS, *READING ACHIEVEMENT, *GRADE 1, BEGINNING READING, BASIC READING, WORD RECOGNITION, BASIC VOCABULARY, VISUAL DISCRIMINATION, UNIVERSITY OF MAINE, ORONO, MAINE, MURPHY DURRELL READING READINESS ANALYSIS, LEE CLARK READING READINESS TEST,

FOUR NULL HYPOTHESES WERE TESTED TO EVALUATE SELECTED READING READINESS TESTS AS PREDICTORS OF FIRST-GRADE READING ACHIEVEMENT. FIVE SCHOOLS IN THE BANGOR, MAINE, CITY SCHOOL SYSTEM WERE RANDOMLY CHOSEN. IN EACH SCHOOL, ONE CLASS USED THE EXPERIMENTAL PROGRAM WHICH EMPHASIZED SUPPLEMENTARY VOCABULARY INSTRUCTION WITH READINESS TRAINING. ANOTHER CLASS USED THE CONVENTIONAL PROGRAM WITHOUT VOCABULARY INSTRUCTION. BOTH PROGRAMS WERE BASED ON THE HARPER-ROW BASAL READERS. READING READINESS WAS DETERMINED BY THE SELECTED READING READINESS TESTS DURING THE SECOND WEEK OF THE SCHOOL YEAR. READING ACHIEVEMENT WAS TESTED BY INDIVIDUAL WORD RECOGNITION TESTS BASED ON THE VOCABULARY OF THE FOUR PREPRIMERS USED AND BY SPACHE'S DIAGNOSTIC READING SCALES. RESULTS STATISTICALLY ANALYZED AT THE UNIVERSITY OF MAINE COMPUTING CENTER INDICATED THAT PREDICTION OF SUCCESS IN FIRST-GRADE READING WAS SIGNIFICANTLY IMPROVED BY USING COMPOSITE SUBTESTS, BY CLEARLY DEFINING THE ACHIEVEMENT CRITERIA, AND BY ORGANIZING AND CONTROLLING INSTRUCTIONAL VARIABLES. NINE SUBTESTS WERE IDENTIFIED, AND IT WAS FOUND THAT INTACT TESTS DID NOT REALLY MEASURE FEATURES OF READINESS. AN EXTENSIVE BIBLIOGRAPHY AND TABULATED TEST RESULTS ARE INCLUDED IN THIS FINAL REPORT OF A STUDY DONE UNDER CONTRACT WITH THE DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE. (NS)

ED015118

PA 24

BR FINAL REPORT
~~Project~~ No. 6-8894
Grant No. OEG-1-7-068894-0323

SELECTED READING READINESS TESTS AS PREDICTORS OF SUCCESS IN READING

July 31, 1967

RE 001 007

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

SELECTED READING READINESS TESTS AS PREDICTORS OF SUCCESS IN READING

Project No. 6-8894
Contract No. OEG-1-7-068894-0323

Robert E. Lowell

July 31, 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

University of Maine

Orono, Maine

RE 001 007

CONTENTS

	Page
Introduction	1
The Problem	1
Procedures	2
Predicting Reading Achievement	6
Limitations of Reports in the Literature	10
Summary	12
Results of the Study	13
Predictive Results	13
The Experimental Effects	20
Factor Analysis Results	22
Summary and Conclusions	23
The Problem	23
Procedures	24
Summary of Results	25
Conclusions	27
References	28
Appendix A	30

LIST OF TABLES

Table		Page
1	Predictors of January Performance on the Harper-Row Test of Word Recognition	13
2	Predictors of May Performance on the Diagnostic Reading Scales - Test of Word Recognition	14
3	Predictors of May Performance on the Diagnostic Reading Scales - Test of Oral Reading	15
4	Predictors of May Performance on the Diagnostic Reading Scales - Test of Silent Reading	16
5	Predictors of May Performance on the Diagnostic Reading Scales - Test of Consonant Sounds	16
6	Predictors of May Performance on the Diagnostic Reading Scales - Test of Vowel Sounds	17
7	Predictors of May Performance on the Diagnostic Reading Scales - Test of Consonant Blends	17
8	Predictors of May Performance on the Diagnostic Reading Scales - Test of Common Syllables	18
9	Predictors of May Performance on the Diagnostic Reading Scales - Test of Blending	18
10	Predictors of May Performance on the Diagnostic Reading Scales - Test of Letter Sounds	19
11	Tests of the Differences Between the Multiple r's Predicting the Ten Reading Achievement Criteria	20
12	Tests of Differences for January and May Reading Achievement	21
13	Factor Matrix for Males	22
14	Factor Matrix for Females	23

Introduction

The literature supports the contention of Barrett[1], and others, that reading readiness tests available today are predictively related to reading achievement, but individually they do not adequately sample the universe of factors identified as being most related to reading achievement. For example, auditory and visual discrimination abilities similar to those required in the act of reading have been revealed to be two highly related factors, yet current reading readiness tests vary widely in measurement of these factors, and a number of the tests do not contain measures of both.

The literature does not reveal how reading readiness tests are related to reading achievement for different stages of the first grade reading program or for differing methods of instruction. The lack of evidence in these two realms seems especially important considering the purposes behind the tests - to direct the reading instructional program and to help the teacher identify areas of readiness to be taught.

Finally, there is only a small amount of information available relative to what reading readiness tests actually measure. Very few studies of readiness tests have included a factor analysis or have attempted to identify the variables measured by the tests. There is little information about the number and kind of factors measured by readiness tests.

Therefore, this study evaluated reading readiness tests as predictors of reading achievement for the different stages of the first grade reading program under two different instructional emphases. This study also analyzed selected part and intact readiness tests to determine a best composite predictor of reading achievement and to identify the factors which readiness tests measure.

The Problem

The problem involved answering the following questions. First, what is the relationship between selected reading readiness tests and reading achievement under two differing instructional programs for two different stages of first grade? Second, what was the influence of a vocabulary supplemented reading readiness program on reading achievement? Third, what are the factors measured by the various reading readiness tests?

---

[1] Thomas C. Barrett. "Predicting Reading Achievement Through Readiness Tests." Reading and Inquiry, vol. 1o, Delaware: International Reading Assoc., 1965. pp. 26-28.

To answer question one and two the following hypotheses were tested using their alternate (null) hypotheses.

Hypothesis 1. The intact readiness tests are significantly related to reading achievement at the end of the preprimer program and at the end of first grade for the control and experimental groups.

Null. The intact readiness tests are not significantly related to reading achievement at the end of the preprimer program and at the end of first grade for the control and experimental groups.

Hypothesis 2. The composite of readiness subtests most highly related to reading achievement are more highly related to reading achievement than any intact readiness test at the end of the preprimer program and at the end of first grade for the control and experimental groups.

Null. The composite of readiness subtests most highly related to reading achievement have the same relationship to reading achievement as the intact readiness tests at the end of the preprimer program and at the end of first grade for the control and experimental groups.

Hypothesis 3. The relationship between readiness and reading achievement is higher in the experimental group at the end of the preprimer program and at the end of first grade.

Null. There is not a significant difference in the relationship between readiness and reading achievement for the control and experimental groups at the end of the preprimer program and at the end of first grade.

Hypothesis 4. Reading achievement at the end of the preprimer program and at the end of first grade is significantly higher in the experimental group.

Null. There is no significant difference in reading achievement for the experimental and control groups at the end of the preprimer program and at the end of first grade.

Question three was answered using a principle-factors, factor analysis of the readiness tests as a means of identifying the number and kind of factors measured by the readiness tests.

Procedures

Reading Readiness Tests. Tests used to measure readiness for reading were selected for study using the following rationale.

Only tests which claimed to measure different features of readiness were chosen because it was the intent of this study to identify factors of readiness and the relationship of those factors to reading achievement.

Several characteristics of the tests which could be classified under administerability were considered. For example, the tests had to have clearly written understandable directions for the teacher, including time limits, number of subjects to test, conditions of testing, and clearly written instructions to be read to the children. The directions for the pupils should be easily understood by the age group being tested. The arrangement of figures on pages, type size, print intensity, and time limits had to be geared to the age level being tested. The tests had to have these characteristics built into the total testing format.

In addition, the manuals had to include a rationale and background for the test's development, the purpose of the test, description of the contents, interpretation of results, suggestions for use of results, and a description of statistical analyses including reliability, validity, and standard error of measure.

Whatever evidence there was in the literature concerning the relative merits of the tests was used to help make selections. A survey of the literature was helpful in determining the tests which were chosen from the many which were available.

The Murphy-Durrell Reading Readiness Analysis (1965 edition), the Visual Test of the Murphy-Durrell Diagnostic Reading Readiness Test (1949 edition), the Lee-Clark Reading Readiness Test (1962 revision), and the Test of General Ability (1959 edition) were chosen for use in this study. This set of tests was selected because it represents a range of measures of several factors which have been shown to be related to reading. The tests are all group measures which in combination were administerable within the time span of one week and which had time requirements per subtests which were suitable for first grade use. They appeared to be as valid and reliable as any available for use in the estimation of readiness for reading in first grade.

Reading Achievement Tests. These tests were selected for many of the same technical reasons given above, especially those involving sound measurement practices. Only individual test of reading achievement were considered in order to maximize the possible achievement of the pupils, and to place the responsibility for testing in the hands of the researcher rather than the classroom teacher. In the study of early development in reading, these were considered important features.

The measurement of reading achievement at the end of the preprimer program posed a problem because very few standardized tests are available for this purpose, and those which could be used are not as reliable as desired, are particularly invalid because little time has elapsed in which reading achievement could take place, and the content of the tests is not similar to the material being learned in reading. Hence, a test was prepared using the vocabulary of the basal series in which the pupils were learning to read and involved conditions closely approximating testing procedures used by first grade teachers

in their instructional program. This test consisted of a list of all the words the child had been taught. The test had content and curricular validity because it was made up from the exact words used in teaching reading and the testing procedures were similar to the informal methods of assessment typically used by first grade teachers. The test was titled, Word Recognition Test, Harper-Row Preprimers (1966).

Reading achievement at the end of first grade was measured by a standardized individual test of reading performance - Diagnostic Reading Scales by George D. Spache (1963). This test was chosen because it appears to measure several dimensions of reading and was suitable in terms of the measurement criteria above. Further, it is especially well suited to use at this grade level because the content of reading matter was graded by a readability formula[2] found to be well suited to primary grade reading levels.

The Sample Population. A random sample of ten classrooms (250 children) from five different schools of the city school system of Bangor, Maine, was selected for this study from the entire system consisting of eleven elementary schools located in various regions of the city.

The sample size was determined according to the types of analyses of data to be performed and to ensure sufficient subjects on whom complete data would be available at the close of the study. The questions of sex differences, treatment effects, and multiple regression of the predictor of variables on the criterion variable require at least a hundred subjects in order that one have reasonable control over sources of error. The relationship between variables is more reliably derived when the comparison group sizes are similar.

A homogeneous sample from one school system was chosen as a control over factors affecting the internal consistency of an experimental setting. In this study the variable of teaching supplementary vocabulary was under examination. More accurate conclusions could be drawn about this variable when the sample included similar teachers, supervisory policies, materials, and methods of instruction. The conditions under which learning of reading took place were considered more alike in one system than would be the case for a sample drawn from several school systems. These were factors favoring internal experimental consistency. Further, a city school system was chosen because the pupils, learning under the homogeneous conditions above, are more representative of populations to which one might wish to generalize conclusions reached as a result of this study. The subjects were all first graders from families and home situations similar to those which can be found in many other communities similar to the city in which the sample of this study was taken.

---

[2] George Spache, "A New Readability Formula," El. School J. 53: 410-13, March 1953.

Administering the Readiness Tests. The battery of reading readiness tests was administered to all subjects in this study during the second week of school in September 1966. Ten teachers from the substitute rolls of the Bangor Superintendent of Schools' office were hired to administer the tests. These examiners were randomly assigned to the classrooms and gave the tests according to a schedule prepared for them. That schedule was made by randomly assigning the tests for administration.

Sending examiners into the classrooms was not considered a reactive situation because this practice in testing is used in the Bangor school system and the subjects of this study have been tested under similar circumstances in the kindergarten.

The tests were scored and checked by a team of clerks hired for that purpose only. Each test was scored twice, once each by two different clerks.

Reading Instruction. Teachers used the readiness testing information, recommendations from the kindergarten teachers, advice of their supervisor, their own readiness checklists and their generalized judgment of pupils' readiness for instruction in reading as bases upon which to organize reading groups.

One classroom in each school operated under experimental conditions and the other functioned as it typically had in the past. Both groups employed the basal reading series published by Harper-Row, Incorporated.

The experimental classrooms were organized so that all children immediately began learning the vocabulary of the first preprimer of the Harper-Row basic reading series. Some groups began in preprimers without any readiness training and others were given readiness training in the reading program of the basal series. However, the children at the readiness level were given supplemental vocabulary instruction beginning in the week following readiness testing. The emphasis here was upon enlarging the reading vocabulary of the children at the readiness level.

The control classrooms had some groups begin instruction in the first preprimer in the week following readiness testing. However, the children who began the readiness program of the Harper-Row series did not receive supplemental vocabulary instruction and hence were delayed in arriving at formal instruction in reading.

These two instructional programs in the Harper-Row basic reading series continued until the end of first grade.

Reading Achievement Testing. Achievement in reading was measured at two different points in the first grade. A measure was taken at the end of the preprimer program and at the end of the first grade.

The preprimer stage of reading was considered complete when the first groups completed the fourth preprimer of the reading series. Achievement at that time was determined by an individual test of word recognition using the vocabulary of the four preprimers of the Harper-Row basic reading series. Two examiners were hired to administer the test and were trained in the testing procedures. A testing schedule was made by randomly assigning the examiners to the classrooms and was accomplished in one week.

The results were summarized by a clerk and reported to the teachers in the week following achievement testing. A summary was also sent to the supervisor of instruction.

Reading achievement at the end of the year was determined by the administration of a standardized individual diagnostic test of reading performance, the Diagnostic Reading Scales, (1963) by George D. Spache.

A team of ten examiners were hired for that purpose. The examiners were given training in the administration and scoring of the test. Each examiner also tested several subjects prior to beginning testing in the classrooms. The examiners did not know if they were testing control or experimental classes.

Data summary sheets were prepared for each classroom and the entire amount of data on subjects was entered at the computer center for statistical analysis utilizing programs of analysis of covariance, stepwise multiple regression, and factor analysis.

Predicting Reading Achievement

In general, the purpose of prediction in education according to Monroe[3] is to "provide information which may be used in the guidance and counselling of individuals." In his review of prognosis as a topic of educational research he indicated that prediction is an integral part of the entire educational structure. The promotion and selection processes used reflect prognosis. By virtue of the fact that prognosis is widely held to be a necessity, accurate prediction becomes imperative. He further stated, "Educational prognosis is essential to the point of determining the nature of the educational program for each child." Modern methods of prediction in education are a substitute for "fortune-telling" techniques of earlier days. Today the preferred modes of prognosis should use scientific methods of data collection and analysis. The body of knowledge derived from scientific methods has replaced the older gross, general, outward appearances which education had to rely upon in times prior to the influences of recent, more sophisticated researchers using appropriate research designs with more cautiously drawn inferences.

---

[3] Walter S. Monroe, Editor. Encyclopedia of Educational Research, Revised Edition. New York: The Macmillan Company, 1956. p. 874.

Goslin[4] offered a view of prognosis consistent with the previous author. His attentions, for the most part, were directed at prediction in general. He particularly emphasized that the ability of a test to predict was dependent upon the relationship between the abilities required in the performance being predicted and those measured by the test. The more alike the predictor and predicted in content and nature, the better the predictive ability of the test. The ideas he presented were compatible with the following statement by Monroe's summary of prediction of reading success:

"The foregoing summary indicated that real progress has been made in understanding the factors and conditions that influence reading readiness and in developing measures that predict success in learning to read. With the facts now available it should be possible to carry on studies in the future that will be very productive in clarifying thinking concerning the requisites for learning to read, in developing tests that will reveal the extent of a pupil's readiness for reading, and in modifying teaching during the prereading period in order to promote with increased effectiveness the types of development that prepare for reading."[5]

In her review of research on reading readiness, Gunderson[6] was in agreement with this statement. She believed there should be three reasons for giving readiness tests, to determine those children who are ready or unready for reading, to identify areas of instruction for reading, to identify areas of instruction to receive attention, and to predict reading achievement. However, the latter of the three uses she isolated and declared unjustified. Apparently she does not consider the first two uses related to prediction, whereas Goslin[7] and Monroe[8] appeared to identify an interrelationship between the

---

[4]David A. Goslin, The Search for Ability. New York: Russell Sage Foundation, 1963. p. 153.

[5]Monroe, op. cit. p.879.

[6]Doris V. Gunderson, Research in Reading Readiness. Department of Health, Education, and Welfare, Bulletin No. 8, 1964, Washington, D.C.: Government Printing Office, 1964. p.32.

[7]Goslin, op. cit. p.153.

[8]Monroe, op. cit. p. 87.

first two uses and the third. One would believe from other sources than Gunderson[9] that using tests as predictors reveals the efficacy of tests as measures of whatever it is that they are measuring. The belief of Gunderson that instruction directed as needs identified by readiness necessarily tests eliminated the predictive value of the tests was inconsistent with the fact that researchers have identified higher and higher relationships between readiness measures and reading achievement. The researchers appeared to be operating under the hypothesis that the more alike the predictor and predicted, the better the prediction; rather than the idea that one should avoid manipulation of variables in order not to disturb the relationship. They behaved as if the prediction could only be true if the course of action indicated by the test results were followed. Such research over the years resulted in improved measures of readiness and modified programs of instruction. Despite the fact that tests have had improved predictive validity, those measures available today are not entirely in agreement in content. Barrett[10] reported that tests disagreed about what was important to measure with the exception of visual discrimination. Even here, though, the tests vary greatly as to what is included as a measured task of visual discrimination. One test may have the child identify the difference between geometric forms while another has the child differentiate similarly spelled words with minute differences in configuration; father, farther. It was Barret's[11] contention that researchers in reading readiness appear to be converging on accurate predictors as the factors of readiness become better identified and as the factors of reading achievement become better identified. He suggested that predictor tests should be selected with specific programs and success criteria in mind.

Several reports were available which show the relationship between readiness tests and reading achievement. In general they reported correlations between .40 and .70. Henig[12] compared the Lee-Clark Readiness Test with teachers forecasts and found both to be significantly related to reading success. The fact that teachers could predict as well as the test negated its use and led him to conclude that teachers be recommended as predictors over the tests. Spaulding[13] reported a correlation of .46 between the Harrison-Stroud Reading Readiness Test and the Gates Primary Reading Test. The Metropolitan

---

[9] Gunderson, op. cit. p. 32.

[10] Thomas C. Barrett, "Predicting Reading Achievement Through Readiness Tests." Reading and Inquiry, Vol. 10. Newark: Delaware: International Reading Assoc., 1965. pp. 26-28.

[11] Barrett, Ibid., p. 28.

[12] Max X. Henig, "Predictive Value of Reading Readiness Tests and of Teachers' Forecasts." Elementary School Journal, 50, Sept. 1949, 41-46.

[13] Geraldine Spaulding, "The Relation Between Performance of Independent School Pupils on the Harrison-Stroud Reading Readiness Tests and Reading Achievement a Year Later." 1955 Fall Testing Program in Independent Schools and Supplementary Studies. Bulletin 67, New York: Educational Records Bureau, February 1956. pp. 73-6.

Readiness Test had a predictive relationship of .59 with the same reading test for a similar population. Certain subtests of each test appeared to be better predictors than others. Karlin[14] and Bremer[15] tested the ability of readiness tests to predict reading success and found readiness tests to be poor predictors. Karlin concluded that research should be undertaken to identify what is being measured by readiness tests. Powell and Parsley[16] reported that the Lee-Clark Reading Readiness Test could reasonably predict group performance in reading. It predicted most accurately for the middle range of youngsters and poorest for the high range of scores on the test.

Other studies of reading tests combined the results of reading testing with various information in an attempt to improve prediction. This appears logical in view of the conclusion of Traxler and Townsend[17] that "if readiness tests have an advantage over teacher estimates, it is that prediction can be obtained on the basis of the tests at the very beginning of first grade or even before the children enter grade one." This statement was made because teachers could judge readiness as well as tests only if they were permitted eight weeks of observation. In their opinion this was a serious loss of time, perhaps causing a delay in reading instruction.

Recent studies supporting the use of readiness measures were carried out by Nash,[18] Barrett,[19] and Dykstra.[20] These researchers

---

[14]Robert Karlin, "The Prediction of Reading Success and Reading Readiness Tests." Elementary English, 35, May 1957, 320-2.

[15]Neville Bremer, "Do Readiness Tests Predict Success in Beginning Reading?" Elementary School Journal, 59, Jan. 1959, 222-4.

[16]Marvin Rowell and Kenneth M. Parsley, Jr., "The Relationships Between First Grade Reading Readiness and Second Grade Reading Achievement." Journal of Educational Research, 54, Feb. 1961, 229-31.

[17]Arthur E. Traxler and Agatha Townsend, Eight More Years of Research In Reading: Summary and Bibliography. Bulletin No. 64. New York: Educational Records Bureau, 1955. p. 12.

[18]Pat Neff Nash, "The Effectiveness of Composite Predictors or Reading Success in the First Grade." Doctor's Thesis. Denton, Texas: North Texas State University, 1963. Dissertation Abstracts. XXIV, 1482-3.

[19]Thomas C. Barrett, "Visual Discrimination Tasks as Predictors of Success in First Grade Reading Achievement." The Reading Teacher, Vol. 18, Jan. 1965, 276-82.

[20]Robert Dykstra, "Auditory Discrimination Abilities and Beginning Reading Achievement." Reading Research Quarterly, Vol. I, No. 3, Spring 1966, pp. 5-34.

utilized composites of measures as predictors. They reported correlations at the upper end of the range typically found in the literature. Barrett,[21] for example, identified a multiple correlation coefficient of .71 between several visual discrimination factors and word recognition for boys and girls combined. Nash[22] reported three different composites of measures with multiple correlations of .74, .73 and .72, respectively. She recommended the lowest of the three composites because it could be utilized by a classroom teacher while the other two groupings necessitated the use of projective testing and sociometric technique.

Influential in the formulation of this study was the early research which tended toward considering single factors as predictors and the recent trend of studies involving composite measures with improved prediction coupled with the conclusions from Goslin[23] and Monroe[24] relative to the similarity of the predictor and predicted variables.

Limitations of the Reports in the Literature

There were the expected weaknesses in research design and statistical analysis that comes from hindsight. However, there was obvious growth in the sophistication of the researchers over the years since the 1920's when reading readiness became a research concern. Several instances of design problems were evident with the most notable flaw being that of arbitrary subject selection and assignment to treatment. Recent studies indicated a greater awareness of the pitfall of this procedure and the strength involved in randomly selected and assigned subjects.

Violations of statistical procedures were reported with the most notable centering around the misapplication of the T-test and extravagant inferences from correlations. The study by Pratt[25] is

---

[21]Barrett, op. cit., p. 281.

[22]Nash, op. cit., p. 1482-3.

[23]Goslin, op. cit., p. 153.

[24]Monroe, op. cit., pp. 874, 879.

[25]Willis E. Pratt, "A Study of the Differences in the Prediction of Reading Success of Kindergarten and Non-Kindergarten Children." Journal of Educational Research, 42, Mar. 1949, 525-33.

illustrative of the abuse of the T-test. He applied the statistic to non-randomly assigned groups of disparate sizes. It is questionable that the differences he reported were really as significant as claimed. It is more likely that the claimed confidence levels were artificially depressed and not really as highly significant as reported. In other instances one researcher would extol the virtue of a correlation of .55 while another was cautious of a higher magnitude. As stated earlier, these conditions are within bounds of expectancy considering the background and preparation of educational researchers. The trend appears to be toward greater sophistication of design and increased skill in application of statistics with greater caution on the inferential level. In the latter regard, caution about the inferences of cause and effect surrounding correlation was issued some time ago by Good, Barr, and Scates.[26]

Regarding the particular weakness of early attention to single predictors, recent investigators attended to multiple factors as predictors. De Hirsch, Jansky, and Langford[27] claimed that the three typical measures of readiness: a. readiness tests, b. intelligence tests, and c. informal evaluations, were all useful, but they had limitations. Readiness tests lacked formulation of education strategies. Intelligence tests failed to account for perceptual functioning at various intelligence levels. Informal evaluations are not readily duplicated and too many teachers lacked training, intuition, or experience in reliable evaluation. Thus, an inherent weakness of previous research has been the measures used. As new measures were developed, continued research was recommended. The weaknesses of readiness tests were pointed out by Barrett[28] who revealed the lack of agreement between contents of tests. Likewise, Heilman[29] summarized the status of readiness tests weaknesses as being attributed to the fact that the tests only included factors which test writers believed were related to reading and some tests included

---

[26] Carter V. Good, A.S. Barr, and Douglas E. Scates, Methoddogy of Educational Research. New York: D. Appleton-Century and Co., Inc., 1936. pp. 559-62.

[27] Katrina De Hirsch, Jeannette Jansky, and William Langford, Predicting Reading Failure - A Preliminary Study. New York: Harper and Row, Publishers, 1966. p. 3.

[28] Thomas C. Barrett, "Predicting Reading Achievement Through Readiness Tests." Reading and Inquiry, Vol. 10, Newark, Delaware: International Reading Assoc., 1965, pp. 26-28.

[29] Arthur W. Heilman, Principles and Practices of Teaching Reading. Columbus, Ohio: Charles E. Merrill Books, Inc., 1961. p. 28.

items more similar to the experiences of some children and thus tended to "overrate" them.

Still another limitation of recent studies of prediction of first grade reading achievement has been a lack of attention to differences which might exist at various points throughout first grade or attributable to instructional emphases at different times of the year or in different materials of instruction during the year. For example, it seems apparent that the instructional emphasis at the beginning of first grade, if using basal readers, involves more visual discrimination than auditory. And, later stages, above pre-primer levels, the emphasis involves more auditory discrimination. The literature did not contain reports of any studies which examined this characteristic. Several studies have been carried out which involved different modes of instruction but no predictions were involved in them.

Another, and what appears extremely important, observation is that very little information is reported in which a factor analysis was performed as a means of attempting to identify what the tests measure. Instead, we were left as Heilman[30] said we were, with the opinions of test authors as the contents of readiness tests. It would appear that studies involving factor analyses of readiness factors are necessary in order to assist in developing better readiness tests.

Summary

A summary of the literature on reading readiness is succinctly supplied in the following statement by Gunderson.[31]

"Results of research conducted in the past have established quite conclusively that readiness for reading is determined by a constellation of factors."

A similar viewpoint on the part of reading authorities was offered by Russell[32] who wrote:

"The modern concept of readiness is that it is based on a combination of physical, mental, social, and psychological factors."

---

[30] Heilman, Ibid., p. 28.

[31] Doris V. Gunderson, Research in Reading at the Primary Grade Level. Office of Education Bulletin 42. Washington: U.S. Government Printing Office, 1963. p. 2.

[32] David H. Russell, Children Learn to Read. New York: Ginn and Co., 1961, p. 168.

It seems a universal finding that readiness is a multi-factored condition which must be determined from bases which account for those factors and their interrelationships. It would appear that research should be recommended which is directed at the solution of the problem of identification of readiness factors and the improved prediction of reading achievement. Such research would necessarily call for careful attention to the measures of prediction, achievement, and their reliability and validity.

Continued study of methods for measuring factors of readiness appeared to be warranted since the literature did not supply a final answer to the question of what readiness is, how it can best be determined, or how well reading achievement can be predicted. In particular, it seemed that a study was needed which would find information to help clarify these points. This study attempted to supply such information by answering the questions stated earlier which were derived from a search of the literature and the researchers' experiences with normally and abnormally developing readers.

RESULTS OF THE STUDY

Predictive Results

A computer program for stepwise multiple regression analysis was used in determining the relationship between the independent (predictor) variables and the dependent (criterion) variables. In that program the computer first identified the independent variable which accounted for the greatest amount of the sums of squares of the dependent variable in question. The computer then located the independent variable contributing the greatest portion of the remaining sums of squares. This procedure repeated until all the independent variables with a minimum value of F = 1.00 were considered in the reduction of the sums of squares. Any variables with F values below 2.41 for the control and 2.44 for the experimental groups were excluded from the ten tables which follow. Thus, all the variables included in the composites were significant predictors at the 1% level.

Table 1. Predictors* of January Performance on the Harper-Row Test of Word Recognition

Control (N=84)	R^{2**}	Experimental (N=96)	R^2
Murphy-Durrell (Test Total)	.434	Murphy-Durrell (Letter Names)	.391
Chronological Age	.066	Lee-Clark (Concepts)	.074
TOGA (Test Total)	.038	Lee-Clark (Letter Symbols)	.036
Lee-Clark (Word Symbols)	.017	TOGA (Test Total)	.017
		Lee-Clark (Test Total)	.056
Total	.555	Total	.594

*In tables 1-10, only predictors significant at the 1% level were included in order of entry into the composite.
**R^2 - the amount of variance estimated by the independent variables.

The composites of measured variables best predicting January reading achievement are presented in Table 1. Here the amounts of estimated variances for the control and experimental group were not significantly different.[1] However, the contents within each composite were different. The variables used to estimate the R^2 of the control group were two intact test totals combining with chronological age and the subtest of the Lee-Clark Readiness Test. The estimators in the experimental composite were three subtests and two intact test totals. The tendency for test totals to predict better for the control group and individual subtests to predict better for the experimental group followed through tables 1-10.

Composites of test scores were better predictors than the intact test totals. A test of the differences between r's for composites and intact batteries revealed a significant difference between the total test r's and the experimental composite r for the Lee-Clark and TOGA.[2] The test of difference between composites and intact test totals for the control group indicates no significant difference between the Murphy-Durrell and the composite multiple r but a significant difference (1% level) between the TOGA total and the multiple r of the composite.

Table 2. Predictors of May Performance in the Diagnostic Reading Scales - Test of Word Recognition

Control (N=84)	R^2	Experimental (N=96)	R^2
Murphy-Durrell (Letter Names)	.281	Lee-Clark (Concepts)	.567
TOGA (Test Total)	.026	Murphy-Durrell (Phonemes)	.052
		Murphy-Durrell (Learning Rate)	.011
		Lee-Clark (Letter Symbols)	.011
		Lee-Clark (Word Symbols)	.311
		Murphy-Durrell (Letter Names)	.006
Total	.307	Total	.958

[1] Table 11, page 80, contains a summary of the tests of differences between experimental and content composite multiple r's predicting the criterion variables.

[2] A complete table of intercorrelations can be found in appendix A.

Table 2 presents the amount of variance estimated by the independent variables at the end of school, May 1967. The magnitude of R^2 in this table are much different from Table 1. The relationship between the control group Word Recognition and the independent variables was considerably different from the amount of estimated variance for the experimental group. The latter prediction was composed entirely of subtests which accounted for 95.8% of the variance of the group on that test while the control group estimate of variance was 30% and included one intact test total and one subtest. The multiple r's for these two composites were significantly different beyond the 1% level.

The Murphy-Durrell Letter Names Test estimated about 28% of the variance for the control group with the Tests of General Ability estimating about 2% of the remaining variance. These two tests were the only ones significantly related to the criterion of May word recognition for the control group. Within that composite the r for the Murphy-Durrell Test Total was not significantly different from the composite r. The TOGA total r was significantly different from the composite r. The structure of the composite for the experimental group was quite different, being made up entirely of six subtests from within intact tests which estimated most of the variance of that group. Each of the subtest r's was significantly different from the composite r.

Table 3. Predictors of May Performance in the Diagnostic Reading Scales - Test of Oral Reading.

Control (N=84)	R^2	Experimental (N=96)	R^2
Murphy-Durrell (Test Total)	.348	Lee-Clark (Concepts)	.291
TOGA (Test Total)	.026	Murphy-Durrell Total	.210
Murphy-Durrell (Learning Rate)	.024	Lee-Clark (Word Symbols)	.281
		Chronological Age	.008
		TOGA (Information)	.005
Total	.398	Total	.795

The first two tables included a similar criterion, word analysis, while table 3 contains information on the dependent variable - oral reading. The pattern of relative amounts of estimated variances per composite which emerged in Table 2 continued into this table and throughout tables 3 through 10. The test of the differences between multiple r's for tables 2-10 revealed the experimental multiple r's to be significantly different from control group multiple r's.

In table 2 the Murphy-Durrell Test Total contributed strongly in both composites with a larger amount of estimated variance under the control program. The remaining significant predictors estimated lesser amounts of variance and completed the two composites. Although the amount of estimated variance for the experimental group was not as high as in Table 2, it nonetheless was significantly different (1% level) from the contrasting figure for the control group on this dependent variable.

A test of the difference between r's for the Murphy-Durrell total and composite, and the TOGA total and composite, within the control group, revealed no significant differences. In contrast, the Murphy-Durrell test total r in the experimental composite was significantly different (1% level) from the composite r.

Table 4. Predictors of May Performance in the Diagnostic Reading Scales - Test of Silent Reading

Control (N=84)	R^2	Experimental (N=96)	R^2
Murphy-Durrell (Test Total)	.407	Lee-Clark Concepts	.317
Lee-Clark (Test Total)	.034	Murphy-Durrell (Test Total)	.177
		Lee-Clark (Test Total)	.273
		Chronological Age	.008
Total	.441	Total	.775

Table 4 was constructed using the results of regression analysis involving the dependent variable of silent reading in May. The trend of significantly better prediction for the experimental group was maintained here with R^2 figures closely approximating the results reported in Table 3. Here, again, the Murphy-Durrell test total emerged as an estimator in both composites with a higher estimate in the control. In the experimental group, the Lee-Clark test total and subtests of Concepts combined with the Murphy-Durrell test total to estimate most of the variance of the composite. The R^2 of the Murphy-Durrell test total of the control group was supplemented by the Lee-Clark test total which added 4% of estimated variance to the estimate. This was the instance in tables 1 through 10 when a single intact test came closest to being the only significant predictor of a criterion.

The Murphy-Durrell test total r was not significantly different from the control group composite r. The other three test total r's of the table were significantly different from their respective composite r's.

Table 5. Predictors of May Performance in the Diagnostic Reading Scales - Test of Consonant Sounds.

Control (N=84)	R^2	Experimental (N=96)	R^2
TOGA (Information)	.115	Lee-Clark (Concepts)	.576
Murphy-Durrell (Phonemes)	.038	Murphy-Durrell (Phonemes)	.049
Lee-Clark (Letter Symbols)	.034	Lee-Clark (Test Total)	.014
		Murphy-Durrell (Learning Rate)	.012
		Lee-Clark (Word Symbols)	.011
		Lee-Clark (Letter Symbols)	.287
		Murphy-Durrell (Letter Names)	.004
		Murphy-Durrell (Visual Test - 1949)	.002
Total	.187	Total	.955

Table 5 reports the relationship between the best composites of independent predictors of the May dependent variable, Consonant Sounds. It can now be seen that the contents of the composites vary with the different criterion. The criterion of this table was more specific and narrowly defined than that of tables 1-4 and called forth more subtests in the composite for the experimental group with a significantly (1% level) higher estimate of variance than for the control group. For the control group there were only three significant variables which combined to estimate 18.7% of the variance. Those three variables were subtests from intact tests. The greatest difference between any control and experimental total R^2 was found in this table. Only a small amount of the control variance could be estimated while 95.3% of the variance of the experimental group was estimated.

The Lee-Clark intact test r was significantly (1% level) different from the composite r in which it was contained.

Table 6. Predictors of May Performance in the Diagnostic Reading Scales - Test of Vowel Sounds.

Control (N=84)	R^2	Experimental (N=96)	R^2
Murphy-Durrell (Test Total)	.271	Lee-Clark (Concepts)	.559
TOGA (Test Total)	?	Murphy-Durrell (Test Total)	.022
		Lee-Clark (Test Total)	.028
		Lee-Clark (Word Symbols)	.230
		Murphy-Durrell (Letter Names)	.008
Total	.293	Total	.847

Table 6 includes the composites of predictors which emerged as the best estimators of the variance of still another narrow and specific reading achievement criterion, vowel sounds. Again, the predictive r was significantly higher for the experimental group where subtests were prevalent in the composite, and the intact test r's were significantly different from the composite r. The Murphy-Durrell test total r for the control group was not significantly different from the composite r. In that same group, however, the Lee-Clark test total was significantly different at the 5% level.

Table 7. Predictors of May Performance in the Diagnostic Reading Scales - Test of Consonant Blends.

Control (N=84)	R^2	Experimental (N=96)	R^2
Murphy-Durrell (Test Total)	.381	Lee-Clark (Concepts)	.557
Murphy-Durrell (1949 ed. visual)	.041	Murphy-Durrell (Test Total)	.085
		Lee-Clark (Test Total)	.014
		Murphy-Durrell (Phonemes)	.009
		Lee-Clark (Word Symbols)	.010
		Lee-Clark (Letter Symbols)	.257
		Chronological Age	.004
Total	.422	Total	.936

Table 7 reveals the relationship between significant predictors of reading achievement and the criterion, common syllables. The Lee-Clark Concepts test continued to contribute first and foremost as in the experimental predictive composites in tables 2 through 7. This trend persisted through table 10. In several instances: tables 2, 5-8, and 10, this subtest entered the composite first and contributed at least 50% of the estimated variance. The two composite r's were significantly different at the 1% level. The Murphy-Durrell and Lee-Clark test total r's and the composite r for the experimental group were significantly different. The other intact test r in the control group was significantly different from its composite r.

Table 8. Predictors of May Performance in the Diagnostic Reading Scales - Test of Common Syllables.

Control (N=84)	R^2	Experimental (N=96)	R^2
Murphy-Durrell (Test Total)	.452	Lee-Clark (Concepts)	.583
TOGA (Information)	.037	Murphy-Durrell (Test Total)	.045
Lee-Clark (Letter Names)	.036	Lee-Clark (Test Total)	.018
Lee-Clark (Test Total)	.022	Lee-Clark (Letter Symbols)	.272
		Chronological Age	.008
		Murphy-Durrell (Phonemes)	.002
Total	.547	Total	.928

A comparison of tables 7 and 8 shows that the better composites in each were similar in terms of order of entering variables, the amount of the variance they estimated, and the total. The experimental composite r was significantly different from the control composite r. In table 8, as in tables 1-4 and 6-10, the Murphy-Durrell test total r contributed the greatest share of the estimated variance in the control group composite and was not significantly different from the control composite r. The other test total r's in the table were not significantly different from the composite r's in which they were contained.

Table 9. Predictors of May Performance in the Diagnostic Reading Scales - Test of Blending.

Control (N=84)	R^2	Experimental (N=96)	R^2
Murphy-Durrell (Test Total)	.142	Lee-Clark (Concepts)	.316
Lee-Clark (Letter Symbols)	.035	Murphy-Durrell (Letter Names)	.074
		Lee-Clark (Test Total)	.030
		Murphy-Durrell (Phonemes)	.015
		Lee-Clark (Word Symbols)	.190
		Chronological Age	.013
Total	.177	Total	.638

In the seven previous tables the experimental composite estimated at least 75% of the variance of the criterion while the regression analysis for the criterion variable of table 9 estimated only 63%. The other patterns of significantly better prediction under experimental circumstances and the intact tests were not being as good estimators as the composites persisted through this table. The Murphy-Durrell Test Total r again was not significantly different from the control group composite r. The only other test total r was significantly different from the r of the composite in which it was contained.

Table 10. Predictors of May Performance in the Diagnostic Reading Scales - Test of Letter Sounds.

Control (N=84)	R^2	Experimental (N=96)	R^2
Murphy-Durrell (Test Total)	.312	Lee-Clark (Concepts)	.640
Murphy-Durrell (1949 ed. visual)	.047	Murphy-Durrell (Phonemes)	.025
Lee-Clark (Concepts)	.041	Lee-Clark (Test Total)	.015
		Lee-Clark (Letter Symbols)	.303
		Murphy-Durrell (Letter Names)	.001
		Chronological Age	.001
		Murphy-Durrell (1949 ed. visual)	.0003
Total	.423	Total	.9853

The last criterion used, letter sounds, was estimated as indicated in table 10. A composite emerged which estimated 98% of the variance of the experimental group, and which was significantly different from the control estimate of 42% of the variance. That composite was made up of several subtests and one test total. By contrast, the control group composite estimated only 42% of the variance and included the Murphy-Durrell test total r as the most prominent estimator. That r was not significantly different from the composite r.

The trends which became evident in table one persisted through to table 10. They were: the intact readiness tests were not as good predictors as any composite, although the Murphy-Durrell test total r was a strong contender in the control group estimates and was not significantly different from the control composite r's. The amount of estimated variance was consistently higher for the experimental composites in tables 1 through 10. And finally, the better predicting experimental composites were composed mostly of subtests and the control composites were composed mainly of test totals.

Table 11. Tests of the Differences Between the Multiple r's Predicting the Ten Reading Achievement Criterion.

Criterion	Experimental (N=96) r	Fishers z_1	Control (N=84) r	z_2	$z_1 - z_2$		
1. Word Recognition (Jan.)	.77	1.02	.74	.95	.07	.46	n.s.
2. Word Recognition (May)	.97	2.09	.55	.62	1.47	9.80	*
3. Oral Reading	.89	1.42	.63	.74	.68	4.53	*
4. Silent Reading	.88	1.38	.67	.81	.57	3.86	*
5. Consonant Sounds	.97	2.09	.43	.46	1.63	10.86	*
6. Vowel Sounds	.92	1.59	.54	.60	.99	6.60	*
7. Consonant Blends	.97	2.09	.66	.79	1.30	8.66	*
8. Common Syllables	.96	1.95	.73	.93	1.02	6.80	*
9. Blending	.79	1.07	.42	.45	.62	4.13	*
10. Letter Sounds	.99	2.65	.65	.78	1.87	12.46	*

* Significant at the 1% level.

Table 11 contains the tests of differences between the composite r's which emerged. The experimental group multiple r's were higher, and were significantly different from the control group multiple r's in criterion 2-10, end-of-year measures. The first criterion, a mid-year measure, revealed non-significantly different composite r's. The information in this table was included in the presentation of the results of prediction in pages 70-81.

The Experimental Effects

The results of the analysis of differences in reading achievement are presented in table 12. An analysis of variance was used to determine treatment effects instead of the planned covariance analysis because there were no differences on the prereadiness measures and the correction for no beginning differences would have been a superfluous analysis. The analysis reported here was tested as a fixed effects model. The contrasts between experimental and control groups, sexes, and the interaction of treatments with sexes were the concern in this part of the study.

The comparisons of treatments effects revealed no significant differences on all but one criterion, Consonant Sounds. In that test the groups differed significantly. Otherwise there were no important differences which could be attributed to a specialized treatment effect designed for this study. The differences between treatment means on the criterion Consonant Sounds was not in the expected direction. That is, the control treatment mean was larger than the experimental and differences favoring the experimental effect were anticipated.

The tests of differences for sexes reveals significant differences on four criteria; January word recognition, May word recognition, consonant blends, and letter sounds. The female means for these four criteria were larger than the male. Girls did significantly better than boys on only these criteria.

Table 12 Tests of Differences for January and May Reading Achievement Results (For complete data see Appendix A)

	F - T e s t		
Criterion	Treatment	Sex	Treatment/Sex
January			
Word Recognition	1.74	6.77**	-1.00
May			
Word Recognition	.40	10.14**	1.96
Oral Reading	.30	3.54	.14
Silent Reading	.40	2.46	1.02
Consonant Sounds	6.35*	2.24	-.12
Vowel Sounds	1.64	.11	.06
Consonant Blends	.18	8.88*	.12
Common Syllables	.47	1.65	2.70
Blending	2.30	.21	.43
Letter Sounds	3.69	10.87**	.23

* Significant @ .05 level
** Significant @ .01 level

The third column of table 12 contains tests of the interaction of treatments with sexes. There were no significant differences found which could be attributed to an interaction between treatments and sex. Despite the few significant differences for sexes, there was no carryover to that effect to the interaction of the treatments by sexes.

The summary of the analysis of reading achievement results indicated the experimental treatment of supplementary vocabulary instruction was an impotent variable; that girls achieved better than boys on some tests; and that there was no significant interaction effect.

Factor Analysis Results

The first question of this study was exploratory in nature and the most appropriate method for deriving an answer was a factor analysis. The question was: What are the factors measured by the readiness tests? Tables 13 and 14 contain the factor analysis of the tests used as predictors in this study. In table 14 it can be seen that factor one contained heavy loadings by all the readiness subtests. The nine different subtests appeared to be measuring the same factor when used with girls. The second factor to emerge was chronological age with the only heavy loading occurring on that variable. The answer to the question of what factors were measured by readiness tests was clear-cut when looking at the data for females. The tests measured one generalized factor of reading readiness.

However, table 13 for males revealed a different pattern. Factor one was loaded heavily by tests 3, 6, 7, 8 and 9, and factor two by tests 2, 4, 5, 10, and Chronological Age. The first factor for males was something which involved the interrelatedness of concepts, phonemes, letter names, learning rate, and information. Factor two was related to Chronological Age, letter symbols, word symbols, visual test (letters and words), and reasoning and appeared as a factor different from the first. Further, the tests of readiness worked differently for boys than girls.

Despite what appeared to be difference in test contents, the readiness tests in general did not measure the number of different factors they purported to measure. The four tests contained a total of nine different subtests which appeared to be measures of different aspects of reading readiness. However, those nine subtests measured one factor for girls and two for boys.

Table 13 FACTOR MATRIX FOR MALES (N=129)

Variable	h^2	Rotated Loadings I	II
1. Chronological Age	.40	-.21	0.60
2. Letter Symbols	.49	.39	-.59
3. Concepts	.43	.65	-.09
4. Word Symbols	.54	.22	-.70
5. Visual Test	.62	.37	-.69
6. Phonemes	.57	.74	-.16
7. Letter Names	.51	.60	-.39
8. Learning Rate	.49	.58	-.37
9. Information	.67	.82	-.05
10. Reasoning	.36	.35	-.49

Table 14 FACTOR MATRIX FOR FEMALES (N=124)			
Variable	h^2	Rotated Loadings I	II
1. Chronological Age	.73	.09	-.85
2. Letter Symbols	.39	.63	.01
3. Concepts	.60	.62	.47
4. Word Symbols	.43	.60	-.28
5. Visual Test	.56	.75	-.05
6. Phonemes	.60	.78	.02
7. Letter Names	.64	.76	-.26
8. Learning Rate	.54	.69	-.24
9. Information	.49	.68	.15
10. Reasoning	.34	.58	.09

SUMMARY AND CONCLUSIONS

The Problem

A review of the literature indicated a number of questions about reading readiness tests and their ability to predict success in reading in first grade. The questions were: What was the relationship between selected reading readiness tests and reading achievement under two different instructional programs for two different stages of first grade? What was the influence of a vocabulary supplemented reading readiness program on reading achievement? What were the factors measured by the various reading readiness tests? Several hypotheses were formulated for testing to supply answers to the questions. Those hypotheses as stated below were considered in numerical order in the discussions in this chapter.

Hypothesis 1. The intact readiness tests were significantly related to reading achievement at the end of the preprimer program and at the end of first grade for the control and experimental groups.

Null. The intact readiness tests were not significantly related to reading achievement at the end of the preprimer program and at the end of first grade for the control and experimental groups.

Hypothesis 2. The composite of readiness subtests most highly related to reading achievement were more highly related to reading achievement than any intact readiness test at the end of the preprimer program and at the end of first grade for the control and experimental groups.

Null. The composite of readiness subtests most highly related to reading achievement had the same relationship to reading achievement as the intact readiness tests at the end of the preprimer program and at the end of first grade for the control and experimental groups.

Hypothesis 3. The relationship between readiness and reading achievement was higher in the experimental group at the end of the preprimer program and at the end of first grade.

Null. There was not a significant difference in the relationship between readiness and reading achievement for the control and experimental groups at the end of the preprimer program and at the end of first grade.

Hypothesis 4. Reading achievement at the end of the preprimer program and at the end of first grade was significantly higher in the experimental group.

Null. There was no significant difference in reading achievement for the experimental and control groups at the end of the preprimer program and at the end of first grade.

Question number three in the problem of this study was answered through a factor analysis of the results of fall reading readiness testing. That analysis identified the number and kind of factors measured by the readiness tests.

Procedures

A set of reading readiness tests composed of the Lee-Clark Reading Readiness Test, 1962 Revision; Murphy-Durrell Diagnostic Reading Readiness Analysis, 1965 Revision; Tests of General Ability, 1959 Edition; and Murphy-Durrell Reading Readiness Analysis, 1949 Edition was chosen for study as predictors of success in first grade reading. A test of word recognitioniin the Harper-Row Basic Reading Program was prepared from the vocabulary of the reading books used in the control and experimental classes and administered as the criterion of January reading achievement. The Diagnostic Reading Scales, 1963 Edition was chosen as the end-of-year reading achievement criteria and was administered in May 1967.

A random sample of ten first grade classes from eleven schools of the Bangor, Maine, school system were chosen for participation as the subjects in the study. One of each pair of classes in the five schools was randomly assigned the experimental program of supplementary vocabulary instruction. In the five classrooms designated experimental, the children who received reading readiness instruction

also were given supplementary vocabulary instruction using the vocabulary from the first preprimer of the Harper-Row Basic Reading Series - the materials of instruction for all classrooms.

The reading readiness tests were administered in September 1966 to 253 first grade children. Immediately following testing instruction began in the ten classes with the experimental subjects all receiving vocabulary lessons even though reading readiness was indicated for several children within each of the five experimental classes. The five control classes were taught as the teachers typically had in the past and did not begin instruction in vocabulary for children on the readiness training level. The instructional programs continued in this pattern throughout the year.

The first measure of reading achievement was administered in January 1967 when the first group in the ten classes completed the fourth preprimer of the reading series and signaled the end of the preprimer program. At that time a word recognition test composed of the vocabulary of the four preprimers was administered.

Reading instruction continued as planned and the remaining achievement criteria, the Diagnostic Reading Scales, were administered in May, 1967. At the time of the final achievement measurement the number of subjects had been reduced to 180. Losses were due to absence at time of testing, withdrawal from school and exclusion of individuals who were repeating grade one.

The testing results for the entire study were correlated and sent to the University of Maine Computing Center where multiple regression, analysis of variance, and a factor analysis were performed. Information from those analyses was used in testing the hypotheses of the study and formulating the following conclusions.

Summary of Results

Hypothesis one was not rejected on the basis of the evidence gathered in the study. Each of the intact tests was a significant predictor for one or more criterion with the Murphy-Durrell Test being the most consistent in its prediction. The reading readiness tests chosen for examination in this study were significantly related to the reading achievement criteria which were used. (See tables 1-10) Thus, the first null hypothesis was rejected and the research hypothesis not rejected.

A test of null hypothesis 2 led to a non-rejection of part of the hypothesis and a rejection of the remainder. That is, the Murphy-Durrell intact test was not significantly different from the composite r for the control group criteria, but the experimental composites were all significantly different from any intact tests contained within them. Likewise other test totals in the control group composites were significantly different from the composite.

The test of null hypothesis 3 revealed a division of the hypotheses similar to that of the preceding. There was a non-rejection of the hypothesis of no differences between the experimental and control composites of prediction. The criterion of achievement in January was predicted equally well by the experimental and control composites. However, the May achievement criteria were all predicted significantly better by the experimental composites.

Thus prediction of reading achievement from composites of readiness subtests was advantageous and prediction was greatly enhanced by the experimental program circumstances despite the lack of significant differences occurring in the tests of the next hypothesis.

Null hypothesis three was rejected on the basis of the evidence in the study. Only the achievement criterion of consonant sounds yielded significant differences with the larger mean attributed to the control treatment. Thus the experimental variable was not potent enough to cause significant differences in reading achievement in January or May.

Despite the lack of significant differences attributable to a treatment, the experimental circumstances apparently were influential in improving the predictive ability of subtests because prediction was much better for the experimental group.

The results of the factor analysis revealed that the readiness tests measured differently for boys than girls. When measuring females, all nine tests measured the same factor but when used on boys the tests took on factorial complexity. For example, the subtests of the Lee-Clark Test measured two factors. The concepts test measured factor one and the remaining subtests measured factor two. This meant that the Lee-Clark Test was uni-factorial for girls and bi-factorial for boys. The Murphy-Durrell 1949 Edition Visual Test measured factor two along with the Reasoning subtest of the TOGA. The Murphy-Durrell Readiness Test (1965) was uni-factorial for both boys and girls. This makes one wonder over the value of such a test when it requires long periods of time to administer and then really does not measure the three different factors it purports to measure. The TOGA, like the Lee-Clark, was bi-factorial for boys but uni-factorial for girls.

It would seem that this evidence should lead one to conclude that the tests which measure different factors are preferable to

those which measure only a single dimension of the constructs which they purport to measure. This data, along with the facts in tables 1-10, should lead one to the conclusion that the readiness tests as they exist are not very good predictors of success in reading but one can gain in efficiency and accuracy by combining a variety of measures. For example the Lee-Clark and TOGA tests provided information about boys in two factors. Either of those two tests would provide information on one factor for girls.

The evidence in tables one through ten for the experimental group predictions lead one to regard the Lee-Clark test quite highly. It predicted well (especially the concept test), was desirable for use with first grade children because it was a short test, and yet it did measure two different factors of readiness when administered to boys. The use of this test over the others in the total battery would constitute a considerable saving of testing time.

It should also be concluded that reading readiness tests are not as good at doing what they claim to do as they could be, and that they do not necessarily measure what they claim to measure unless one identifies how the tests work under varying learning circumstances.

Conclusions

On the basis of the evidence gathered in this study it was concluded that prediction of success in first grade reading using reading readiness tests was significantly improved by using composites of subtests, clearly defining the achievement criteria, and organizing and controlling the teacher variables even though the variables involved may not significantly improve reading achievement.

It was also concluded that intact readiness tests by themselves were not as good predictors as desired. The fact that the intact tests were measures of only one factor for girls and two for boys indicated the inefficiency of using intact tests. The differences between sexes in factors measured and in achievement led to the conclusion that the readiness measures worked differently for boys than girls.

It was further concluded that readiness tests did not measure discrete features of readiness and were not very good measures to use as indicators of success unless the instructional program was identified and the achievement goals well defined. The reading readiness tests used in this study were measures of nine different traits of readiness, yet the tests were apparently not different enough to discriminate and become identified as measures of separate factors in a factor analysis.

References

Barrett, Thomas C. "Predicting Reading Achievement Through Readiness Tests." Reading and Inquiry, Vol. 10, 1965, 26-8.

Barrett, Thomas C. "Visual Discrimination Tasks as Predictors of Success in First Grade Reading Achievement." The Reading Teacher, Vol. 18, Jan. 1965, 276-82.

Bremer, Neville. "Do Readiness Tests Predict Success in Beginning Reading?" Elementary School Journal, 59, Jan. 1959, 222-4.

DeHirsch, Katrina; Jansky, Jeannette; and Langford, William. Predicting Reading Failure - A Preliminary Study. New York: Harper and Row, Publishers, 1966. 144 p.

Dykstra, Robert. "Auditory Discrimination Abilities and Beginning Reading Achievement." Reading Research Quarterly, Vol. I, No. 3., Spring 1966, pp. 5-34.

Good, Carter V.; Barr, A.S.; and Scates, Douglas E. Methodology of Educational Research. New York: D. Appleton-Century Co., Inc., 1936. 676 p.

Goslin, David A. The Search for Ability. New York: Russell Sage Foundation, 1963. 204 p.

Gunderson, Doris V. Research in Reading at the Primary Level. Office of Education Bulletin 42. Washington: U.S. Government Printing Office, 1963. 114 p.

Heilman, Arthur W. Principles and Practices of Teaching Reading. Columbus, Ohio: Charles E. Merrill Books, Inc., 1961. 465 p.

Henig, Max X. "Predictive Value of Reading Readiness Tests and Teacher's Forecasts." Elementary School Journal, 50 Sept. 1949, 41-46.

Karlin, Robert. "The Prediction of Reading Success and Reading Readiness Tests." Elementary English, 35 May 1957, 320-2.

Monroe, Walter S., Editor. Encyclopedia of Educational Research, Revised Edition. New York: The Macmillan Company, 1956. 1520 p.

Nash, Pat Neff. The Effectiveness of Composite Predictors of Reading Success in First Grade. Doctor's Thesis. Denton, Texas: North Texas: North Texas State University, 1963. Dissertation Abstracts: XXIV,1483-4.

Powell, Marvin, and Parsley, Kenneth M., Jr., "The Relationships Between First Grade Reading Readiness and Second Grade Reading Achievement." Journal of Educational Research, 54, Feb. 1961, 229-31.

Pratt, Willis E. "A Study of the Differences in the Prediction of Reading Success of Kindergarten and Non-Kindergarten Children." Journal of Educational Research, 42, Mar. 1949, 525-33.

Russell, David H. Children Learn to Read. New York: Ginn and Co., 1961, 169 p.

Spache, George D. Reading in the Elementary School. Boston: Allyn and Bacon, Inc., 1964. 356 p.

Spaulding, Geraldine. "The Relation Between Performance of Independent School Pupils on the Harrison-Stroud Reading Readiness Tests and Reading Achievement a Year Later." 1955 Fall Testing Program in Independent Schools and Supplementary Studies. Bulletin 67. New York: Educational Records Bureau, February 1956. pp. 73-6.

Traxler, Arthur E., and Townsend, Agatha. Eight More Years of Research in Reading: Summary and Bibliography. Bulletin No. 64. New York: Educational Records Bureau, 1955. 2834.

Appendix A

Analysis of Variance for the Reading Achievement Criteria

January - Word Recognition

Source	DF	SS	MS	F	
Treatment	1	632.17	632.17	1.74	n.s.
Sex	1	2458.06	2458.06	6.77	.01
Treatment/Sex	1	-36.54	-36.54	-1.00	n.s.
Error	176	63884.65	362.98		
Total	179	66938.34			

Trt. Means	Sex Means
57.22 Exp.	62.59 Female
60.97	55.19

May - Word Recognition

Source	DF	SS	MS	F	
Treatment	1	.11	.11	.40	n.s.
Sex	1	2.74	2.74	10.14	.01
Treatment/Sex	1	.53	.53	1.96	n.s.
Error	176	48.57	.27		
Total	179	51.97			

Trt. Means	Sex Means
2.03 Exp.	2.177 Female
2.08	1.930

May - Oral Reading

Source	DF	SS	MS	F	
Treatment	1	.38	.38	.30	n.s.
Sex	1	4.40	4.40	3.54	n.s.
Treatment/Sex	1	.18	.18	.14	n.s.
Error	176	219.27	1.24		
Total	179	224.24			

Trt. Means	Sex Means
2.34 Exp.	2.44 Female
2.24	2.13

May - Silent Reading

Source	DF	SS	MS	F	
Treatment	1	.46	.46	.40	n.s.
Sex	1	2.78	2.78	2.46	n.s.
Treatment/Sex	1	1.16	1.16	1.02	n.s.
Error	176	200.63	1.13		
Total	179	205.05			

Trt. Means	Sex Means
2.24 Exp.	2.30 ~~Exp.~~ Female
2.13	2.06

May - Consonant Sounds

Source	DF	SS	MS	F	
Treatment	1	11.51	11.51	6.35	.05
Sex	1	4.07	4.07	2.24	n.s.
Treatment/Sex	1	-.22	-.22	-.12	n.s.
Error	176	320.15	320.15	1.81	
Total	179	335.52			

Trt. Means	Sex Means
2.75 Exp.	3.15 ~~Exp.~~ Female
3.26	2.84

May - Vowel Sounds

Source	DF	SS	MS	F	
Treatment	1	4.04	4.04	1.64	n.s.
Sex	1	.27	.27	.11	n.s.
Treatment/Sex	1	.15	.15	.06	n.s.
Error	176	432.01	2.45		
Total	179	436.49			

Trt. Means	Sex Means
.84 Exp.	1.03 ~~Exp.~~ Female
1.14	.95

May - Consonant Blends

Source	DF	SS	MS	F	
Treatment	1	.44	.44	.18	n.s.
Sex	1	20.88	20.88	8.88	.01
Treatment/Sex	1	.30	.30	.12	n.s.
Error	176	412.63	2.34		
Total	179	434.20			

Trt. Means	Sex Means
2.68 Exp.	3.05 ~~Exp.~~ Female
2.78	2.37

May - Common Syllables

Source	DF	SS	MS	F	
Treatment	1	1.16	1.16	.47	n.s.
Sex	1	4.05	4.05	1.65	n.s.
Treatment/Sex	1	6.59	6.59	2.70	n.s.
Error	176	429.71	2.44		
Total	179	441.53			

Trt. Means	Sex Means
1.81 Exp.	2.03 ~~Exp.~~ Female
1.97	1.73

May - Blending

Source	DF	SS	MS	F	
Treatment	1	5.21	5.21	2.30	n.s.
Sex	1	.48	.48	.21	n.s.
Treatment/Sex	1	.98	.98	.43	n.s.
Error	176	398.64	2.26		
Total	179	405.32			

Trt. Means	Sex Means
1.50 Exp.	1.38 ~~Exp.~~ Female
1.16	1.28

May - Letter Sounds

Source	DF	SS	MS	F	
Treatment	1	3.47	3.47	3.69	n.s.
Sex	1	10.22	10.22	10.87	.01
Treatment/Sex	1	.23	.23	.23	n.s.
Error	176	166.75	.94		
Total	179	180.69			

Trt. Means	Sex Means
3.29 Exp.	3.65 ~~Exp.~~ Female
3.57	3.18